The Estimate of Human Capital from two Sets of Observed Indicators: Formative and Reflective

Giorgio Vittadini, University of Bicocca Milan, Italy **Pietro Giorgio Lovaglio,** University of Bicocca Milan, Italy.
Giorgio Vittadini, Department of Statistics, University of Bicocca-Milan,Via Bicocca degli Arcimboldi 8, 20126 MILAN, ITALY. (E-MAIL giorgio.vittadini@unimib.it)

KEY WORDS: Human Capital,
Latent Variable, Pls, Lisrel,
Actuarial approach

1. Introduction

Dagum and Vittadini (1996) stressed that, from the statistical point of view, Human Capital (HC) can be expressed as a Latent Variable (LV). Taking into account the prospective definition of HC, it can be defined as the flow of earned income throughout an individual's life span. They specified and estimated a general recursive model (Dagum, 1994; Dagum et al., 2003) purporting to explain the determination and the distribution of income, net and gross wealth, debt and HC.
In a recent proposal, Vittadini, Dagum, Lovaglio and Costa (2003) combine a zerodimensional LV approach and an actuarial mathematical approach to estimate the HC of American Households in monetary units, putting greater emphasis on economic theory.
This is because HC involves both its investment amounts in families (formative indicators) and its effect on income (reflexive indicators).
The importance of HC for the determination and the accumulation of income justifies its extension to a multiple dimension.
The aim of the present paper is to generalize previous approaches: HC is considered as a LV of dimension one that contributes to explaining Household Income as a unique dependent variable, proposing the case of HC in dimension two; in this case HC is composed of two LVs: the Educational dimension (EduHC) and Working Experience dimension (JobHC), underlying the process of determination of earned income and capital income, connected to observed indicators in accordance with economic theory.
The methodology estimated the HC of Italian Households utilizing a survey by Banca d'Italia for 2000. The Survey on Household Income and Wealth (Banca d'Italia, 2002) covers 8,001 households composed of 22,268 individuals and 13,814 income-earners

2. The Economic Model

We can define HC as a bidimensional LV (Educational Human Capital, Working Experience HC) underlying the process of determination and of accumulation of Earned Income and Capital (Property) Income.
As in previous papers, the observed indicators of HC can be partitioned in two blocks: indicators describing the investment in HC and indicators describing the return of HC in a life span (reflective indicators).
The indicators which take into account the factors regarding the formation of HC (defined formative indicators) are divided into four groups: indicators involving information concerning Household, Head (H), Spouse (S), Parents of H and Parents of S. The formative indicators are Household Net Wealth, decomposed into three terms: real assets (WREAL), financial assets (WFIN), total debt (DEBT), and two blocks regarding the block of Educational variables and the block of Working Experience as investment in HC.
More specifically, the variables of educational investment in HC (**Z**) contains WFIN, WREAL, DEBT and the variables in the $\mathbf{Z_1}$ block, where H and S stand for Head and Spouse of the Household Head, respectively.
The investment in HC (**X**) derived from working experience is described by WREAL, WFIN, DEBT and by the variables in the $\mathbf{X_1}$ block shown in Table 1.
The life span return on HC is described by the reflective indicators regarding income decomposed into $\mathbf{y_1}$ and $\mathbf{y_2}$, representing the Net Disposable Labour Income and Net Disposable Capital Income, respectively.
In particular $\mathbf{y_1}$ stands for Household Earned Income (Compensation of employees + Pensions and net transfers + Economic assistance + Income from self-employment), and $\mathbf{y_2}$ for Household Capital Income (Income from real assets + Income from financial assets).

Educational Indicators Z =(Z_1, WREAL, WFIN, DEBT)
HSCOLAR (SSCOLAR) = H (S)Years of schooling HSEX = H Sex, HLAUREA (SLAUREA)= H (S) Type of Degree (5= Engineering and Medicine, 4=Economics, 3= Humanistic 2=Law and Political Science, 1=Science and Math, 0=no degree), HETA=H age (years), AREA5= geographical area (1=North-east, 2= North-west, 3=Centre, 4=South, 5=Islands), NCOMP= Household Size, CHILD= Household Child, NPERC= Household income earners, HSTUP= H father's Educational Level, HSTUM= H mother's Educational Level, SSTUP= S father's Educational Level, SSTUM= S mother's Educational Level, WREAL=Household real wealth, WFIN=Household financial real wealth, DEBT= Household total debts.

Working Indicators X = (X_1, WREAL, WFIN, DEBT)
HETALAV= H age of entrance in the labour market, HACONTRIB = H number of years of full time job, HETA= H age, HQUAL (SQUAL)= H (S) employment status (1= blue-collar worker, 2=office worker or school teacher, 3=cadre or manager, 4= sole proprietor/member of the arts or professions, 5=other self-employed, 6=pensioner, 7=other not-employed), HSETT (SSETT) =H (S) branch of activity (1=agriculture, 2=industry, 3=public administration, 4=other, 5= not employed), SCONP= S father's job status, HCONP = H father's job status, SCONM= S mother's job status, HCONM= H mother's job status, WREAL=Household real wealth, WFIN=Household financial real wealth, DEBT= Household total debts.

Table 1 Indicators for HC (Banca d'Italia Survey, 2002)

A Path Diagram outlines the described model (figure1)

3. The Nature of LVs: Previous Proposals

In statistical literature there are two different meanings for the term LV. With reference to the LISREL equations, specified in (1),

$$\eta = \mathbf{B}\eta + \mathbf{G}\xi + \gamma \tag{1a}$$
$$\mathbf{y} = \Lambda_y \eta + \Delta_y \tag{1b}$$
$$\mathbf{x} = \Lambda_x \xi + \Delta_x \tag{1c}$$

the model is composed by a structural model (1.a) and by two factor models (1.b) and (1.c), showing how each LV η (endogenous) and ξ (exogenous) is linked to its observed indicators specified in vector form **y** and **x**, respectively. In a general structural model, LV solutions are "true LVs" in the sense of Bentler's definition (Bentler, 1982): "A necessary and sufficient condition for a linear equation system to be a latent variable model is that the dimensionality of the space spanned by the independent variable (i.e., the rank of the covariance matrix) is greater than the dimensionality of the space spanned by the manifest variables" (Bentler, 1982). The LVs can be latent causes or latent effects of their observed indicators; in particular, in the Lisrel Model, the LVs are latent causes.

In another approach, the LVs can be defined as "Unobservable Component Variables" of their observed indicators and can be obtained as their linear combinations. In this case the dimensionality of the spaces spanned by the LVs and the observed indicators is equal (Tenenhaus, 1995; Kmenta, 1991). The partial Least Squares Method (PLS) obtains HC as a linear combination of its indicators by means of an iterative algorithm.

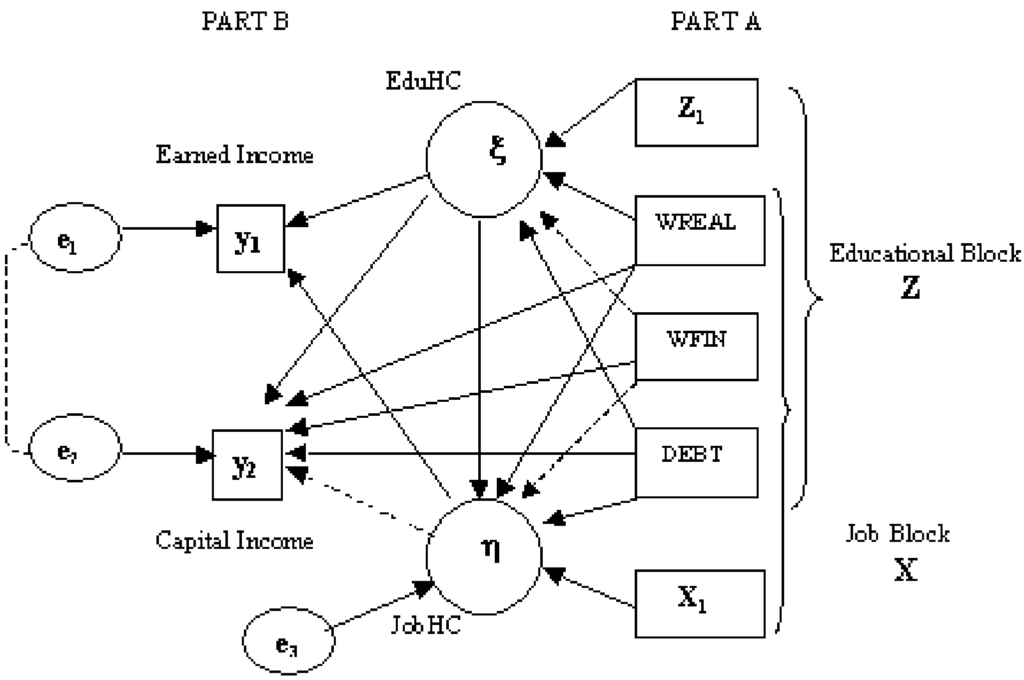

Fig 1: Path Diagram of the Measurement model and Structural model for bidimensional HC (dotted lines in Part A refers to nonsignificant indicators for the contribution of the LV (HC) estimation (Measurement model) and dotted lines in Part B refers to nonsignificant parameters in the simultaneous equations (Structural model).

From the statistical point of view, the major drawback of the Lisrel Model is the problem of the lack of unique solutions (Reiersol, 1950; Guttmann, 1955; Anderson and Rubin, 1956; Lawley and Maxwell, 1963; Joreskog, 1967; Schonemann and Wang, 1972; Schonemann and Steiger, 1978; Steiger, 1979; Schonemann and Haagen, 1987; Vittadini, 1989). Therefore, it can be demonstrated that even if the model is identified, the latent scores are indeterminate. Moreover, there are infinite sets of latent scores for the same identified model.

Differently from the LISREL Model, PLS provides firstly, proxies of LVs as linear combinations of their manifest variables (MVs), and secondly, estimates of the causal parameters and errors.

Proxies of LVs, indicated by η_j^+, j=1,...,m, and ξ_γ^+, γ=1,...,r) are obtained in four steps, which are briefly described in the following sequence.

At the first step, the "outer estimations" of the j-th endogenous LV η_j and γ-th exogenous LV ξ_γ are obtained as linear combinations of observed indicators specified in the matrices $\mathbf{Y}_j$ and $\mathbf{X}_\gamma$ respectively, given arbitrary weights $\beta_j^{(1)}$ and $\alpha_\gamma^{(1)}$ respectively:

$$\eta_j^{+(1)} = \mathbf{Y}_j\, \beta_j^{(1)} \qquad (j=1,\dots,m) \qquad (2a)$$
$$\xi_\gamma^{+(1)} = \mathbf{X}_\gamma\, \alpha_\gamma^{(1)} \qquad (\gamma=1,\dots,r). \qquad (2b)$$

The second step provides the "inner estimation" of η_j and ξ_γ as a linear combination of the "outer estimates" of their adjacent LVs respectively:

$$\eta_j^{+(2)} = \mathbf{\Psi}_j^{+(1)}\, \upsilon_j \qquad (j=1,\dots m) \qquad (3a)$$

$$\xi_\gamma^{+(2)} = \mathbf{\Phi}_\gamma^{+(1)}\, \tau_\gamma \qquad (\gamma=1,\dots r), \qquad (3b)$$

where $\mathbf{\Psi}_j^{+(1)}$ contains the outer estimations of the exogenous and endogenous LVs which are connected to η_j, $\mathbf{\Phi}_\gamma^{+(1)}$ contains the outer LVs estimations connected to ξ_γ; υ_j and τ_γ are the vectors of correlation coefficients, or regression coefficients (or their signs) between $\eta_j^{+(1)}$ and $\mathbf{\Psi}_j^{+(1)}$ and between $\xi_\gamma^{+(1)}$ and $\mathbf{\Phi}_\gamma^{+(1)}$, respectively (Wold, 1982).

At the third step, each component β_{jv} of the weight vector β_j in (2a) is estimated by means of a regression of $\mathbf{y}_{jv}$ on $\eta_j^{+(2)}$:

$$\mathbf{y}_{jv} = \eta_j^{+(2)}\, \beta_{jv}^{(2)} + \zeta_{jv} \quad (v=1,\dots,k_j;\ j=1,\dots,m), \qquad (4)$$

and the weights α_γ in (2b) are obtained by means of multiple regression of $\xi_\gamma^{+(2)}$ on $\mathbf{X}_\gamma$ (Tenenhaus, 1995):

$$\xi_\gamma^{+(2)} = \mathbf{X}_\gamma\, \alpha_\gamma^{(2)} + \omega_\gamma \quad (\gamma=1,\dots,r) \qquad (5)$$

At the fourth step, the weights obtained in (4) and (5) are used to compute new "outer estimates" of η_j and ξ_γ:

$$\eta_j^{+(3)} = \mathbf{Y}_j\, \beta_j^{(2)} \qquad (6)$$
$$\xi_\gamma^{+(3)} = \mathbf{X}_\gamma\, \alpha_\gamma^{(2)}. \qquad (7)$$

By iterating equations (2)-(7) the algorithm converges to weights β_j and α_γ and consequently definitive "outer estimates" η_j^+ and ξ_γ^+ are achieved. From these estimates, the PLS provides the estimation of parameter and error matrices by means of simple/multiple regressions, as indicated in Wold (1982).

Thanks to the explicit, albeit approximate, estimation of the LVs, no identification problem arises in the PLS approach. Analogously, no indeterminacy problem arises, because the dimensionality of the spaces spanned by observable variables (the rank of the observable variables covariance matrix) is equal to the rank of the covariance matrix spanned by the LVs.

In the Path Diagram of figure 1 the unidimensional LVs η and ξ, in accordance with economic theory regarding the process of HC formation, are considered as being compromises between "true" LVs and Unknwon Component Variables. In effect, they are "Unknown Component Variables" of formative indicators while they are latent causes of reflective indicators as economic theory affirms regarding HC. Therefore in this case, neither the Lisrel Model or PLS are adequate for estimating HC, from both an economic and statistical point of view.

Applying the PLS algorithm to the model specified in Figure 1, we can exemplify the procedure of estimating the latent scores following the scheme in Figure 2.

In step 2 of the PLS algorithm, ξ is updated by a linear combination of $\mathbf{Y}$, with $\mathbf{Y}=(\mathbf{y_1}, \mathbf{y_2})$, and $\eta^{+(1)}$ obtained in the first step ($\eta^{+(1)}=\mathbf{X}\beta^{(1)}$), while η is updated as a linear combination of $\mathbf{Y}$ and $\xi^{+(1)}$ (obtained in the first step $\xi^{+(1)}=\mathbf{Z}\alpha^{(1)}$):

$$\xi^{+(2)} = \mathbf{Y}\,\tau_1 + \eta^{+(1)}\,\tau_2 \text{ and } \eta^{+(2)} = \mathbf{Y}\,\upsilon_1 + \xi^{+(1)}\,\upsilon_2 \qquad (8)$$

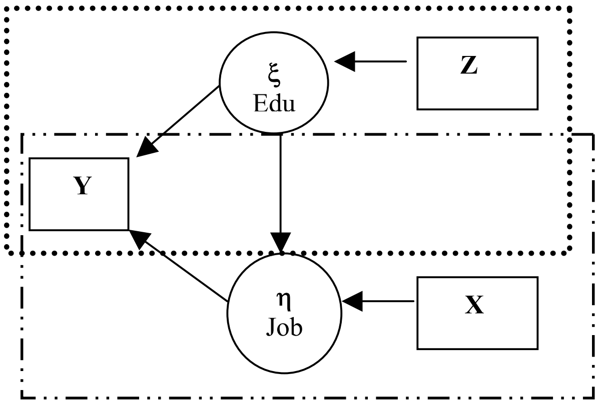

Fig.2 PLS algorithm for the estimate of η and ξ

In this way each LV is obtained as a linear combination of its reflective indicators (**Y**) and of formative indicators of the other LV.

Consequently, the PLS solution does not respect the causal links between observed indicators and LVs in the model, because in the economic model proposed in Figure1, ξ is caused by **Z** and causes **Y,** η is caused by **X** and causes Y.

Moreover, in the second step of the PLS algorithm, the scores of η are used to update the scores of ξ (in the inner estimation step). Nevertheless in economic theory ξ causes η.

4. A Revised Model

From the logical point of view, Lisrel and PLS are not adequate for analysing this problem, therefore to overcome their limitations a more consistent model is proposed.

The HC LVs are defined as " compromise" LVs; true LVs with respect to the reflective indicators earned income and capital income, and "Unknown Composite Variables" with respect to the formative indicators . In this approach the uniqueness problems are overcome and economic causal links are respected.The Structural Model is shown in equations (9) while equations (10)-(11) show the measurement models for ξ and η:

$$\mathbf{Y} = \xi\lambda + \eta\mathbf{b} + \Delta_1 \qquad \Delta_1 \sim (\mathbf{0}, \Sigma) \tag{9}$$

$$\eta = \xi\, c + \mathbf{Xk} \tag{10}$$

$$\xi = \mathbf{Z\,g} \text{ with } \mathbf{g'Z'Zg}=1 \tag{11}$$

where the LVs ξ and η are hypothesired as being rank one (n*1), $\mathbf{Y}=(\mathbf{y_1}, \mathbf{y_2})$ and Δ_1 (errors in equations) are n*2 matrices, $\mathbf{Z}$ (n*p) and $\mathbf{X}$ (n*m) matrices of observed variables supposed of full rank, λ and $\mathbf{b}$ (1*2) vectors of regression parameters of each $\mathbf{Y}$ column on ξ and η respectively, $\mathbf{g}$ (p*1) vector that define the ξ scores, c and $\mathbf{k}$ (m*1) parameters that assess the weight of ξ and $\mathbf{X}$ to the η score, Σ the error matrix (2*2) of the endogenous variables.

The measurement models for ξ and η in equation (10)-(11) are measured without errors because otherwise the reduced form of the model (9)-(10) would be overidentified.

In fact, ignoring the hypothesis that the regarding indicators errors of indicators are mutually independent (Σ diagonal), like in the present case, "nothing is gained by retaining a disturbance in the causal equation error", because it is would be impossible to empirically distinguish whether the partial correlation was attributable to common distrubance or inherent to correlations among the disturbances (Hauser and Goldberger, 1971).

De Leeuw also affirms that "models along the usual structural equation lines can have problems of identification": therefore he opts for "ignoring the errors" and for choosing the Reduced Rank Regression or Redundancy Analysis (de Leeuw, 2003).

In effect, substituting (10) in (9), we obtain:

$$\mathbf{Y} = \xi\lambda + (\xi\, c + \mathbf{Xk})\,\mathbf{b} + \Delta_1 \tag{12}$$

and substituting (11) in (12) we obtain:

$$\mathbf{Y}=\mathbf{Zg}\,\lambda+(\mathbf{Zg}\,c+\mathbf{Xk})\mathbf{b} + \Delta_1 = \mathbf{Zg_1}\,\lambda_1 + \mathbf{Xk_1}+\Delta_1 \tag{13}$$

where $\mathbf{g_1}$, λ_1 and $\mathbf{k_1}$ are counterparts of $\mathbf{g}$, λ, and $\mathbf{k}$ of equation (9), (10) and (11) in the reduced form.

Equation (13) with the constraints $\mathbf{g_1'Z'Zg_1}=1$ generalizes a Reduced Rank Regression model (Tso, 1981) with rank equal to one for ξ. The reduced form shows that the weights defining ξ can be achieved in equation (13) by projecting $\mathbf{Z}$ on the $\mathbf{Y}$ space. However, following the Path Model and the equation (13), we must project $\mathbf{Z}$ onto the $\mathbf{Y}$ space, net to the indirect by of η (and hence orthogonal to the space spanned by $\mathbf{X}$). In this way we must project $\mathbf{Z}$ onto $\mathbf{Q_X}$ $\mathbf{Y}$, where $\mathbf{Q_X}$ is a n*n matrix defining the orthogonal projector to the $\mathbf{X}$ space. Nevertheless because $\mathbf{X}$ and $\mathbf{Z}$ share $\mathbf{W}$ (WREAL, WFIN and DEBT) as common covariates, we need to project $\mathbf{Z}$ onto $\mathbf{Q_{X1}Y}$, subspace of $\mathbf{Y}$ orthogonal to $\mathbf{X_1}=\mathbf{X}\text{-}\mathbf{W}$ space. At this point to achieve the weights defining ξ we specify the following model:

$$\mathbf{Q_{X1}Y} = \mathbf{Zg_1}\,\lambda_1 + \Delta_1 \qquad \Delta_1 \sim (\mathbf{0}, \Sigma) \tag{14}$$

where $\mathbf{Q_{X1}Y}$ and $\mathbf{Z}$ are (n*2) and (n*p) data matrices, $\mathbf{g}_1$ is a p*1 vector of weights parameters that defines the scores of ξ ($\xi^*=\mathbf{Zg_1}^*$), λ_1 is the (1*2) vector containing the regression coefficients of each $\mathbf{Q_{X1}Y}$ column onto ξ^* and Δ_1 is the n*2 matrix of the residuals in equation (9).

The estimation problem of the RRR model is reduced to the minimization of the trace and determinant of the Gram matrix of the residuals (Tso, 1981).

The estimate of λ_1 conditional to the estimate of $\mathbf{g}_1$ is obvious, under normalized constraint of equation (11):

$$\lambda_1 \,|\, \mathbf{g_1} = \mathbf{g_1}'\,\mathbf{Z}'\,\mathbf{Q_{X1}Y} \tag{15}$$

Secondly, with Σ matrix initially estimated by the ITSUR estimator (Srivastava and Giles, 1987), $\mathbf{g}_1^*$ is obtained as suggested in Tso (1981).

We define $\mathbf{s_1}$ as the first column of the coefficients matrix $\mathbf{S}$ defining the decomposition of $\mathbf{Z}$ ($\mathbf{Z} = \mathbf{US}$ where $\mathbf{U}$ has orthonormal columns and $\mathbf{S}$ is invertible) and $\mathbf{f}_1$ in the first column of matrix $\mathbf{F}$ formed by the eigenvectors of the correlation matrix $\mathbf{RR'}$, with $\mathbf{R} = \mathbf{U'V}$, and $\mathbf{V}$ is the counterpart of $\mathbf{U}$ in the matrix decomposition of $\mathbf{Q_{X1}}$ $\mathbf{Y}$ ($\mathbf{Q_{X1}Y} = \mathbf{VT}$).

Because the canonical variates of $\mathbf{Z}$ are $\mathbf{S^{-1}P}$, with $\mathbf{P}$ the singular vectors of $\mathbf{R}$, and because it is demonstrated that the eigenvectors $\mathbf{F}$ of $\mathbf{RR'}$ coincide with those of the left hand singular vectors ($\mathbf{F}$) of $\mathbf{R}$, we have:

$$\mathbf{g}_1^* = \mathbf{f}_1 / \mathbf{s_1} \tag{16}$$

Briefly, $\mathbf{g}_1^*$ contains the weights defining the first canonical variate of $\mathbf{Z}$ in a canonical correlation framework between $\mathbf{Q_{X1}}$ $\mathbf{Y}$ and $\mathbf{Z}$.

Therefore, we conclude that the estimate of the latent scores ξ^* is a linear combination of $\mathbf{Z}$ variables ($\mathbf{Zg}_1^*$),

but not a linear combination of $\mathbf{Y}$, because ξ is a latent cause of $\mathbf{Q}_{\mathbf{X1}}\mathbf{Y}$ and not vice-versa.
Nevertheless, the weights $\mathbf{g}_1^*$ are obtained by maximising the sum of the coefficients of determination between the components of $\mathbf{Q}_{\mathbf{X1}}$ $\mathbf{Y}$ and $\mathbf{Z}$.
After having obtained ξ^* for the second LV η of the Path Diagram in Figure 1, introducing the value of equation (10) in equation (9) we obtain:

$$\mathbf{Y} = \xi^* \lambda_2 + \mathbf{X}\, \mathbf{k}_2\, \mathbf{b}_2 + \Delta_2 \qquad \Delta_2 \sim (0, \Sigma) \qquad (17)$$

where ξ^* is the estimate by RRR via equation (16) and λ_2, $\mathbf{k}_2$, $\mathbf{b}_2$, Δ_2 are the parameters of the reduced form for η.
By the same arguments explained for ξ, we express $\mathbf{Q}_{\xi^*}\mathbf{Y}$ as dependent variables by projecting the $\mathbf{X}$ space onto the space of $\mathbf{Y}$ orthogonal to ξ^*

$$\begin{gathered}\mathbf{Q}_{\xi^*}\, \mathbf{Y} = \mathbf{X}\, \mathbf{k}_2 \mathbf{b}_2 + \Delta_2 \\ \mathbf{k}_2' \mathbf{X}' \mathbf{X}\, \mathbf{k}_2 = 1 \quad \Delta_2 \sim (\mathbf{0}, \Sigma)\end{gathered} \qquad (18)$$

ξ and η (following the structural equations specified in the Path Diagram) have been estimated as the best linear combination (of rank one) of their formative indicators that best fit their reflective indicators, it is now possible to estimate the causal model shown in Figure 1 as a system of three simultaneous linear equations.

5. The Results

By using the method proposed in section 4 we obtain the result of the model described in Fig.2. Table 2 shows the significant parameters in the educational block, in the job block and finally in the structural model.
The presence of qualitative variables poses no problem because they can be expressed as dummy variables in the estimation procedure.
In Table 2 the most significative parameters (Variable), their statistical significance (Pr>F), and the associated F test (F) are shown
For Educational HC (EduHC) the significant parameters are: years of schooling of H (HSCOLAR), and S (SSCOLAR), real wealth (WREAL), geographical area (AREA5) number of children (CHILD). For the investment in HC derived from work experience (JobHC) we observe a high statistical significance for years of full time job for H, (HACONTRIB), real wealth (WREAL), years of entrance in the labour market for H (HETALAV), age of H (HETA) and interaction of occupation and labour sector (QUAL*SETT) for H and S.
Previous results demonstrate that the contribution of the years of schooling of the Spouse (SSCOLAR) is more important that the contribution of the Head (HSCOLAR) to the formation of educational Household Human Capital.
Moreover, Genetic assets in terms of the Job status of the father of H (HCONP) and S (SCONP) enter the linear combination that defines JobHC, while the years of study of H father (HSTUP) and S parents (SSTUP* SSTUM) contribute to the formation of EduHC.

	Variable	F	Pr > F
EduHC	HSCOLAR	60.39	<.0001
	HLAUREA	5.27	<.0001
	SSCOLAR	82.51	<.0001
	SLAUREA	5.16	<.0001
	HSTUP	3.05	0.0094
	SSTUP*SSTUM	1.78	0.0047
	AREA5	26.02	<.0001
	HSEX	7.85	0.0051
	CHILD	11.5	0.0007
	WREAL	40.42	<.0001
	DEBT	5.2	0.0226
JobHC	HACONTRB	96.46	<.0001
	HETALAV	56.72	<.0001
	HQUAL*HSETT	12.82	<.0001
	SQUAL*SSETT	28.89	<.0001
	SCONP	6.23	<.0001
	HCONP	3.69	0.0003
	HETA	27.53	<.0001
	WFIN	3.08	0.0793
	WREAL	74.39	<.0001
	DEBT	8.43	0.0037

Table 2 Significance of EduHC and JobHC indicators

Finally, in accordance with economic theory, the signs of the coefficients defining EduHC (not reported here) show the important role of the geographical area (AREA5) to the formation of Household HC (North versus South, North versus Central), the positive role of the Real Wealth (WREAL) and the negative role for the Total Debt (DEBT).
As a surprising result we can note the non-significant contribution of financial wealth (WFIN) to the formation of Educational HC (EduHC) as well as a low significance for the formation of working experience HC (JobHC).
For the formation of JobHC, the signs of the coefficients (not reported here) show the (positive) impact of the Head age (HETA) and of the years of full time job, the important (negative) role of the age of entrance to the labour market (HETALAV) and the role (positive) of Real Wealth (WREAL).
The kind of occupation and the labour sector, that enter in the model as interactions (QUAL*SETT), show a major contribution by the Spouse to the formation of JobHC.
At this point, the structural equations, show the causal links between the EduHC, JobHC, $\mathbf{y}_1$ and $\mathbf{y}_2$.
A correlation (ρ) is hypothesized between the errors of $\mathbf{y}_1$ and $\mathbf{y}_2$. **N**evertheless because the test has not refused the hypothesis of null correlation for ρ, we can consider the structural model to be a recursive model with no problems of identification.

The standardized regression parameter estimates in three equations are shown in equation (19), (20) and (21). The t-values are shown in parentheses under the covariates while the R-squared for the equations are shown are shown in parentheses under the error terms

$$\begin{array}{llll} \mathbf{JobHC} = 0.6890\ \mathbf{EduHC} & + \mathbf{e_3} & & (19) \\ \quad (50.57) & (R^2=0.474) & & \end{array}$$

$$\begin{array}{lllll} \mathbf{y_1} = 0.4720\ \mathbf{JobHC} & + 0.2826\ \mathbf{EduHC} & + \mathbf{e_1} & & (20) \\ \quad (25.44) & \quad (15.22) & (R^2=0.486) & & \end{array}$$

$$\begin{array}{lllll} \mathbf{y_2} = 0.1507\ \mathbf{EduHC} & - 0.4119\ \mathbf{DEBT} & +0.5322\ \mathbf{WFIN} + & & \\ \quad (12.78) & \quad (-30.43) & \quad (40.85) & & \\ & + 0.5277\ \mathbf{WREAL} & + \mathbf{e_2} & & (21) \\ & \quad (36.43) & (R^2=0.669) & & \end{array}$$

Equation (19) shows the significant contribution of educational HC (EduHC) to the formation of job HC, while in equation (20) the return of HC in terms of earned income strongly depends on the (positive) contribution of JobHC and EduHC (where the role of the working experience is twice the role of the educational HC). From equations (19) and (20) we observe that JobHC is more important than EduHC in the process of generating Earned Income, but JobHC strongly depends on the educational HC.

Finally the net disposable capital income (**y_2**) can be adequately described as a function of net Wealth in its different aspects (Debt, Real and Financial Wealth); Capital Income in particular strongly depends on Wealth and on total DEBT, less than Educational Human capital; another surprising result is that JobHC does not influence **y_2**.

The goodness of fit index (Joreskog and Sorbom, 1985) based on the residuals (GFI=0.9939), and the GFI adjusted for degrees of freedom (AGFI= 0.9755) for the structural model composed by three equations simultaneously considered shows an excellent goodness of fit.

Finally, Fig.2 and Fig.3 demonstrate the standardized scores of Educational Human Capital (EduHC) and the Working Experience Human Capital (JobHC) respectively.

The approach should be extended to the estimation of Human Capital in monetary units for both HC dimensions in order to estimate personal HC distribution In this way we will have HC as a basic source of information for the implementation of income, wealth and HC redistribution policies.

The results of the monetary approach obtained for the HC component obtained with this method are still in progress.

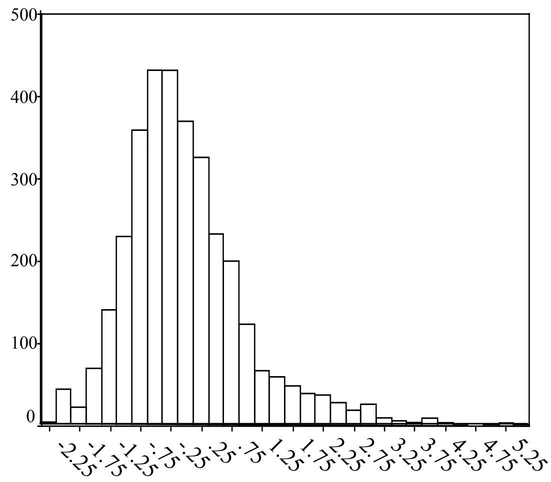

Fig.2: Distribution of EduHC (standardized)

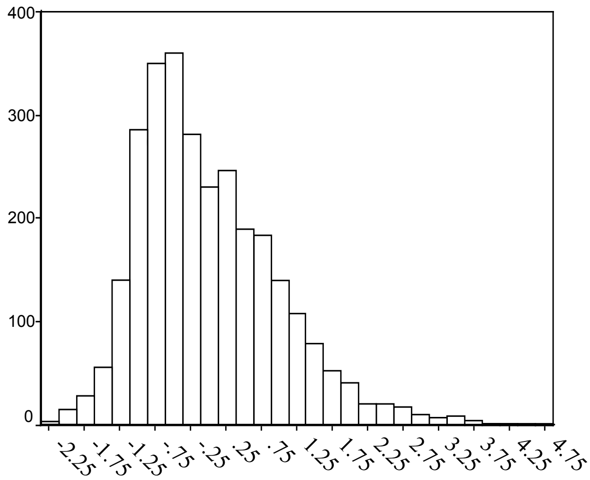

Fig.3: Distribution of JobHC (standardized)

References

Anderson, T. W. and Rubin, M. (1956). Statistical Inference in Factor Analysis. *Proceeding of the third Berkeley Symposium on Mathematical Statistics and Probability* 5, pp.11-150.

Banca d'Italia, (2002). I bilanci delle famiglie italiane nell'anno 2000 (Italian Family Budget, 2000), Supplemento al Bollettino Statistico.

Bekker, P.A. (1989). Identification in Restricted Factor Models, and the Evaluation of Rank Conditions, *Journal of Econometrics* 41, pp.5-16.

Bekker, P.A. and Pollock D.S.G. (1986). Identification of Linear Stochastic Models with Covariance Restrictions, *Journal of Econometrics* 31, pp.179-208.

Bentler, P.M. (1982). Linear System with Multiple Levels and Types of Latent Variables. In *System under Indirect Observation*, Jöreskog K. and Wold H. (eds.), North Holland, Amsterdam, pp. 101-130.

Dagum, C. (1994). Human Capital, Income and Wealth Distribution Models and Their Applications to the USA, *Proceedings of the 154th Meeting of the American,* pp 253-258.

Dagum, C. and Slottje D. J., (2000). A New Method to Estimate the Level of and Distribution of Household Human Capital with Application, *Structural Change and Economic Dynamics*, 11, pp. 67-94.

Dagum, C. and Vittadini, G. (1996). Human Capital Measurement and Distribution, *Proceedings of the 156th Meeting of the American Statistical Association*, Business and Economic Statistics Section, pp. 194-199.

Dagum, C., Vittadini, G., Lovaglio, P. and Costa, M. (2003). A Multiequational Recursive Model of Human Capital, Income and Wealth of Households with Application, *200,3 Proceedings of the American Statistical Association*, Business and Economic Statistics Section [CD-ROM], Alexandria, VA: American Statistical Association.

de Leeuw, J. (2003). Least squares optimal scaling of partially observed linear system, technical report n°360, UCLA University.

de Leeuw, J. and Van Rijckevorsel, J. (1980). Homals and Princals, Some Generalizations of Components Analysis, *Data Analysis and Informatics*, pp. 231-241.

de Leeuw, J. and Young, F. (1978). The Principal Component of Mixed Measurement Level Multivariate Data: an Alternating Least Squares Method with Optimal Scaling Features, *Psychometrika*, 43, pp. 279-281.

de Leeuw, J., Young, F. and Takane, Y. (1976). Regression with Qualitative and Quantitative Variables: an Alternating Least Squares with Optimal Scaling Features, *Psychometrika*, 41, pp. 505-529.

Eisner, R. (1985). The Total Incomes System of Accounts, *Survey of Current Business*, January.

Everitt, B.S. (1984). An Introduction to Latent Variable Models, Chapman and Hall, N.Y.

Fisher, F. (1976). *The Identification Problem In Economics*, Huntingdon, NY: Krieger Publ.

Fuller W.A. (1987). *Measurement Error Models*. Wiley, New York.

Garthwaite, P.H. (1994). An Interpretation of Partial Least Squares, *Journal of American Statistical Association*, 89, pp. 122-127.

Geraci, V. (1976). Identification of Simultaneous Equation Models with Measurement Error, *Journal of Econometrics* 4, 263-283.

Guttmann, L. (1955). The Determinacy of Factor Scores Matrices with Implications for Five other Basic Problems of Common Factor Theory, *British Journal of Statistical Psychology*, 8, pp. 65-81.

Hauser, R.M., and Goldberger, A.S. (1971). The Treatment of Unobservable Variables in Path Analysis, in H.L. Costner (eds.), *Sociological Methodology,* San Francisco: Jossey-Bass, pp. 81-117.

Joreskog, K. (1967). Some Contributions to Maximum Llikelihood Factor Analysis, *Psychometrika*, 32, pp. 443-482.

Jöreskog, K. G. (1973). A General Method for Estimating a Linear Structural Equation System, in Goldeberger A. S. and Duncan O. D. (Eds.), *Structural Eequation Models in the Social Sciences,* Seminar Press, New York, 85-112.

Jöreskog, K. G. (1981). Analysis of Covariance Structures, *Scandinavian Journal of Statistics*, 8, 65-91.

Jöreskog, K. G. (1982). The Lisrel Approach to Causal Model Building in the Social Sciences. In K. G. Jöreskog, & H. Wold (Eds.). Systems under Indirect Observation, *Causality Structure Prediction Vol. 4* (pp. 81-99). New York: North Holland.

Joreskog, K.G. and Sorbom, D. (1985). *LISREL VI; Analysis of Linear Structural Relationships by Maximum Likelihood, Instrumental Variables, and Least Squares*, Uppsala: University of Uppsala.

Lawley, D.N and Maxwell, A. E. (1963). *Factor Analysis as a Statistical Method*. London: Buttherworths.

Lovaglio, P.G. (2001). The Estimate of Latent Outcomes in *Proceedings on Processes and Statistical Methods of Evaluation*, Scientific Meeting of Italian Statistic Society, Rome, Tirrenia, pp. 393-396.

Lovaglio, P.G (2003). The Estimate of Customer Satisfaction in a Reduced Rank Regression Framework, *Total Quality Management*, 16, pp 33-44.

Quiroga, A. (1991). Studies of the Polychoric Correlation and another Measures for Ordinal Variables. *PhD Dissertation*, Uppsala University, Department of Statistics.

Reiersol, O. (1950). On the Identifiability of Parameters in Thurstone's Multiple Factor Analysis. *Psychometrika*, 15, pp.121-149.

Reiersol, O. (1950). On the Identificability of Parameters in Thurstone's Multiple Factor Analysis, *Psychometrika* 15 121-149.

Rothenberg, T.J, (1971). Identification in Parametric Models, *Econometrica*, Vol. 39 pp. 577-591.

Schonemann, P. and Steiger, J.H (1978). On the Validity of in Determinate Factor Scores. *Bulletin of the Psychonomic Society*, 12, pp. 287- 290.

Schonemann, P. and Haagen, K. (1987). On the Use of Factor Scores for Prediction, *Biometrical Journal* 29, pp.835-847.

Schonemann, P. and Wang, M. M. (1972). Some New Results on Factor Indeterminacy, *Psychometrika* 37, pp. 61-91.

Shapiro, A. (1985). Identifiability of Factor Analysis: Some Results and Open Problems, *Linear Algebra and its Application*, 70, pp.1-7.

Srivastava, V.K., Giles D.E.A. (1987). *Seemingly Unrelated Regression Equations Models*, Mercel Dekker, New York.

Steiger, J. H. (1979). The Relationship between External Variables and Common Factors, *Psychometrika,* 44 pp. 93-97.

Tenenhaus, M. (1995). *La Régression PLS: Théorie et Pratique*, Editions Technip, Paris.

Tso, M.K.S (1981). Reduced Rank Regression and Canonical Analysis, *Journal of the Royal Statistical Society*, Series B, 43, pp.183-189.

Vittadini, G. (1999). Analysis of Qualitative Variables in Structural Models with Unique Solutions. In: M. Vichi, O. Opitz, (eds.), *Classifications and Data Analysis-Theory and Applications*, Springer and Verlag, pp. 203-210.

Vittadini, G., Dagum, C., Lovaglio, P.G. and Costa, M. (2003). "A Method for the Estimation of the Distribution of Human Capital from Sample Surveys on Income and Wealth", *American Statistical Association, Business and Economic Statistics Section,* San Francisco, August 3-7, 2003, in JSM Section Proceedings on CD-ROM.

Vittadini, G. and Lovaglio, P.G. (2001). The Estimate of Latent Variables in a Structural Model: an Alternative Approach to PLS. In *PLS and Related Methods*. *Proceedings of the PLS International Symposium*. CISIA CERESTA, Montreuil, France, 423-434.

Wold, H. (1982). Soft Modelling: the Basic Design and some Extension. In: K. Joreskog and Wold H. (eds.), *System under Indirect Observation*, North Holland, Amsterdam, pp. 1-53.

Young, F. (1981). Quantitative Analysis of Qualitative Data, *Psychometrika*, 46, pp. 357-38.